

OPTICO-MECHANICAL ANALOGY: AN AXIOMATIC APPROACH

Alexander Afriat

ABSTRACT: An axiomatic characterization of a ‘two-level Hamiltonian structure’ is proposed, which expresses the optico-mechanical analogy by representing optics and mechanics as (disjoint) classes of models satisfying the axioms. There is the ‘Hamilton-Jacobi level,’ which involves a differential manifold on which the characteristic function satisfying the Hamilton-Jacobi equation is defined; and the ‘symplectic level,’ involving the Hamiltonian, defined on the cotangent bundle of the manifold. The two levels, with the (analogous) structures on them, concern both optics and mechanics.

1 Introduction

The optico-mechanical analogy, which arguably had a central role in the development of wave mechanics,¹ is usually expressed² by describing analytical mechanics and geometrical optics separately, then pointing out a handful of similarities, sometimes by means of a ‘dictionary’ indicating correspondences between analogous objects: between the mechanical potential and the optical index of refraction, between mechanical trajectories and optical rays and so forth. In such approaches the ‘analogy’ has a tenuous, insubstantial status, entirely subordinate to the theories, which are primary. To give the analogy itself more substance and an appropriate primacy, to turn it into an object in its own right (and “give to airy nothing a local habitation and a name”), one can, rather than just noting a few resemblances, abstract a common structure³ possessing optical and mechanical models. The difference is perhaps one of emphasis and degree, for even a statement of similarities might be viewed as a rough description of structure; but the degree seems considerable.

Of the various resources available for the characterization of structure, those of set-theoretical axiomatization, associated chiefly with Patrick Suppes, seem appropriate here and will be adopted. The axioms will not be intended—unlike those of, say, a group or a vector space—as a ‘necessary and sufficient’ basis for deduction, only as a summary of structure, a delineation of *Gestalt*. An appropriate ‘completeness’ (perhaps allowing the derivation of all important theorems) and ‘independence’ are sometimes expected of axioms, but will not be sought here; genuine completeness is unattainable, practically at any rate, and independence is both hard to characterize exactly and unnecessarily stringent. The axioms are only meant to emphasize certain features of the structure that best represent it and are judged significant.

¹ See de Broglie (1925) and Schrödinger (1926, 1934); but also Wessel (1979).

² See Arnol’d (1997) for instance.

³ “Structure,” in the language of model theory, is often a synonym of “model.” Here it is more like a synonym of “theory.”

2 Set-theoretical axiomatization

Set-theoretical methods have been explicitly used for the axiomatization of scientific theories since McKinsey *et al.* (1953) did so, giving rise to a considerable ‘structuralist’⁴ literature. Patrick Suppes, who sought to extend the methods of Bourbaki to the natural sciences, has had a fundamental role in the programme, especially in its early days; subsequent developments, stressing semantics, have also been due to Sneed (1971), Stegmüller (1973, 1980), Balzer *et al.* (1987) and others. Mechanical theories have served as paradigmatic examples since the beginning: ‘classical particle mechanics’ was axiomatized in McKinsey *et al.* (1953), McKinsey and Suppes (1953) and Suppes (1999, 2002); Lagrangian and Hamiltonian mechanics in Sneed (1971). Optical theories have received less attention. Set-theoretical characterization of Hamilton-Jacobi theory or the optico-mechanical analogy has not yet been attempted.

Economy is central to the approach; set-theoretical methods lend themselves to the omission of what is taken for granted or ‘understood,’ such as rules of inference, set theory itself, or staple mathematics. No attempt is made at exhaustive characterization of logical structure; only what is considered relevant and worth mentioning is included. Even the simplest scientific theories can involve many sets, functions and axioms, many more than deserve attention; most are left out.

Suppose we want to give a standard formalization of elementary probability theory. On the one hand, if we follow the standard approach, we need axioms about sets to make sense even of talk about the joint occurrence of two events, for their events are represented as sets and their joint occurrence as their set-theoretical intersection. On the other hand, we need axioms about the real numbers as well, for we shall want to talk about the numerical probabilities of events. Finally, after stating a group of axioms on sets, and another group on the real numbers, we are in a position to state the axioms that belong just to probability theory as it is usually conceived. In this welter of axioms, those proper to probability theory can easily be lost sight of. More important, it is senseless and uninteresting continually to repeat these general axioms on sets and on numbers whenever we consider formalizing a scientific theory. No one does it, and for good reason.⁵

A theory is characterized in the structuralist approach by a set-theoretical predicate—like ‘is a classical particle mechanics’ or ‘is a group’—involving the basic objects and axioms of the theory; or equivalently by the extension of the predicate, in other words by the class of models satisfying the axioms.

⁴ “Sneed, auf den die hier vorgetragene ... Auffassung zurückgeht, hatte sein methodisches Vorgehen ursprünglich als ‘non-statement view’ bezeichnet. ... Bar-Hillel ... hatte ... den Vorschlag gemacht, diese unschöne negative Kennzeichnung durch eine positive zu ersetzen und dafür die Bezeichnung “strukturealistische Auffassung von Theorien”, kurz: “Strukturalismus”, zu verwenden” (Stegmüller 1980 p.2)

⁵ Suppes (2002) p.27.

Here a theory with very different models is axiomatized. Set-theoretical methods free the mathematical theory from association with particular kinds of models, and can be applied, unlike other formal methods, without regard for specific application.⁶ Furthermore a given set-theoretical structure can not only admit diverse physical models, but even abstract mathematical models:

It is one of the theses of this book that there is no theoretical way of drawing a sharp distinction between a piece of pure mathematics and a piece of theoretical science. The set-theoretical definitions of the theory of mechanics, the theory of thermodynamics, and a theory of learning, to give three rather disparate examples, are on all fours with the definitions of the purely mathematical theories of groups, rings, fields, etc. From a philosophical standpoint there is no sharp distinction between pure and applied mathematics, in spite of much talk to the contrary.⁷

Indeed it will be unclear, and unimportant, whether mathematics or physics is really at issue here.

What is meant here by ‘optico-mechanical analogy’ is that two disjoint classes of models—optical and mechanical—satisfy the axioms proposed. The optical models only add semantics to the abstract scheme, whereas further logical structure (basic objects and axioms) is provided for mechanical models.

The general outlines of the structure common to analytical mechanics and geometrical optics will be described with broad brushstrokes, without aiming at generality or rigour. Coordinates will be avoided to keep the presentation simple and clean, not for rigour. Global issues—conjugate points *etc.*—will also be avoided, however relevant they may be to optics.

3 Two-level Hamiltonian structure

The theory expressing the optico-mechanical analogy will itself involve a further analogy, between two ‘substructures’ it contains, one on an underlying manifold, the other on its cotangent bundle. At either level, briefly, the differential of a real-valued function (the characteristic function on the underlying manifold, the Hamiltonian on the cotangent bundle) gets turned into a vector field tangent to a congruence by means of a twice-covariant tensor (the metric tensor in one case, the symplectic two-form in the other); one can say the function *generates* the congruence. Of course there are differences: the symplectic two-form is skew-symmetric whereas the metric tensor is symmetric; the Hamiltonian is largely arbitrary, whereas the characteristic function has to satisfy a differential equation (but with much freedom in the choice of initial conditions). One can speak, perhaps loosely, of a ‘Hamilton-Jacobi’ substructure on

⁶ “... the Suppes approach ... frees us from the prejudice (implicit in the formal language approach) that the formalized theory must have just one big intended interpretation” (Moulines and Sneed 1979 p.74)

⁷ Suppes (2002) p.33.

the underlying manifold and of a ‘symplectic’ substructure on the cotangent bundle. Together they make a ‘two-level Hamiltonian structure.’

Both substructures involve the six objects $M, f_M, df_M, \omega_M^\#, X_M, \Phi_M$: a manifold M , a real-valued function f_M defined on it, the differential df_M , and a twice-covariant tensor $\omega_M^\#$ used to turn df_M into the vector field $X_M = (df_M)^\# = \omega_M^\#(df_M)$ tangent to the congruence Φ_M . The generic manifold or ‘index’ M can have two values, Q and Γ , where

$$\Gamma = T^*Q = \bigcup_{q \in Q} T_q^*Q$$

is the cotangent bundle of the manifold Q , and T_q^*Q the cotangent space dual to the space T_qQ tangent at q to Q . For $M = \Gamma$ we have the symplectic level, where f_Γ is the Hamiltonian $H: \Gamma \rightarrow \mathbb{R}$ and $\omega_\Gamma^\#$ the symplectic two-form $\Omega^\#$ (naturally induced by the cotangent bundle structure), which at a point z of Γ is the transformation $\Omega^\#(z): T_z^*\Gamma \rightarrow T_z\Gamma$ from the cotangent space $T_z^*\Gamma$ to the tangent space $T_z\Gamma$. Setting $M = Q$ gives the Hamilton-Jacobi level, where f_Q is the characteristic function $W: Q \rightarrow \mathbb{R}$, $\omega_Q^\#$ the metric tensor $g^\#: T_q^*Q \rightarrow T_qQ$ (all $q \in Q$), df_Q the covector $p = dW \in T_q^*Q$, and X_Q the vector $\dot{q} = p^\# = g^\#(dW) \in T_qQ$.

The two substructures ‘communicate.’ The projection $\pi(\Phi_\Gamma)$ of the congruence Φ_Γ onto the underlying manifold Q is a crisscrossing tangle, not another congruence. Hamilton-Jacobi theory will be viewed here as a method of extracting a congruence Φ_Q on Q out of the tangle by means of an arbitrary ‘initial’ surface $Q_0 \subset Q$ (of codimension one in Q), together with the *Hamilton-Jacobi equation* $H(q, dW) = \text{constant}$ —which provides a further connection between the two levels, by relating a function $H = f_\Gamma$ on Γ with the differential of a function $W = f_Q$ on Q . Physically Φ_Γ will represent the ‘fictitious’ set of all possibilities—obtained varying over *all* initial conditions—compatible with the Hamiltonian; and Φ_Q the smaller set of all possibilities compatible with the Hamiltonian and with a particular class (selected by Q_0) of initial conditions.

The two substructures form the core of the two-level Hamiltonian structure. Of the aforementioned six objects we will choose three as basic: the function f_M (as the ‘generator’ or ‘input’), the twice-covariant tensor $\omega_M^\#$ (as the fundamental transformation), and the congruence Φ_M (as the ‘output’)—the differential df_M and vector field X_M being too closely related to f_M and Φ_M to be worth including. Rather than the index M we will take its two values, Q and Γ , and also the initial surface $Q_0 \subset Q$. Though basic, the objects $\mathfrak{H} = (Q, \Gamma, f_M, \omega_M^\#, \Phi_M, Q_0)$ are not independent; the three freely-chosen ones are the manifold Q , which determines Γ ; the function f_Γ , which then generates the congruence Φ_Γ ; and finally the surface Q_0 , which gives f_Q and Φ_Q .

As axioms we adopt:

Axiom 1: $\Gamma = T^*Q$ is the cotangent bundle of the n -dimensional differential manifold Q .

Axiom 2: $f_M : M \rightarrow \mathbb{R}$, where $M = Q, \Gamma$.

Axiom 3: The vector field $\omega_M^\#(df_M)$ is tangent to the congruence Φ_M on M , where $\omega_M^\#$ is a twice-covariant tensor field, $\mu \mapsto \omega_M^\#(\mu) : T_\mu^*M \rightarrow T_\mu M$, $\forall \mu \in M$.

Axiom 4: The function $f_Q : Q \rightarrow \mathbb{R} : q \mapsto f_Q(q)$ with $(n-1)$ -dimensional level surface $Q_0 \subset Q$ satisfies equation $f_\Gamma(q, df_Q) = \text{constant}$.

The extension of the predicate ‘is a two-level Hamiltonian structure’ is the class of all systems \mathfrak{H} satisfying Axioms 1-4.

A few words on how the initial surface Q_0 and condition $H(q, dW) = \text{constant}$ determine W , under the (typically mechanical) simplifying assumption that the Hamiltonian is of the form $H = \langle p, p^\# \rangle / 2 + V(q)$, where $\langle x, y \rangle$ is the value of the covector x at the vector y . A distance and a direction will tell us how to advance from a point q_0 of Q_0 : the distance is provided by the condition $\langle p, p^\# \rangle / 2 + V(q_0) = \text{constant}$, the direction by Q_0 , in fact by the $(n-1)$ -dimensional space $T_{q_0}Q_0 \subset T_{q_0}Q$ tangent to Q_0 at q_0 . Calling the constant η , condition $\langle p, p^\# \rangle = 2[\eta - V(q_0)]$ determines a sphere

$$\Sigma_{q_0}^\eta = \{p \in T_{q_0}^*Q_0 : \langle p, p^\# \rangle = 2[\eta - V(q_0)]\}$$

of radius $\sqrt{2[\eta - V(q_0)]}$, which is the required distance. The hyperplane $T_{q_0}Q_0$ determines a direction $\alpha_0 = (T_{q_0}Q_0)^\perp$ ‘normal’ to it, where

$$(T_{q_0}Q_0)^\perp = \{p \in T_{q_0}^*Q : \langle p, \dot{q} \rangle = 0, \dot{q} \in T_{q_0}Q_0\}$$

is the annihilator of $T_{q_0}Q_0$; $\dim \alpha_0 = \dim T_{q_0}Q - \dim T_{q_0}Q_0 = 1$. Of the two points in $\alpha_0 \cap \Sigma_{q_0}^\eta$, we choose the one, say p_0 , that lies in the direction we want. And a momentum $p_0 = dW_0$ for every $q_0 \in Q_0$ is enough, together with the congruence Φ_Γ , to determine the differential dW , the vector field $\dot{q} = g^\#(dW)$ tangent to Φ_Q , and W up to an additive constant. For the unique curve $\gamma_\Gamma : [a, b] \rightarrow \Gamma : t \mapsto (q(t), p(t))$ of Φ_Γ passing through $z_0 = (q_0, dW_0) \in \Gamma$ assigns a differential $dW(t) = p(t)$ to every point $q(t) = \gamma_Q(t)$ in the image $\bar{\gamma}_Q \subset Q$ of the projection $\pi(\gamma_\Gamma) = \gamma_Q : [a, b] \rightarrow Q$ of γ_Γ onto Q . Doing the same for the rest of Q_0 we cover Q and obtain Φ_Q .

4 Optics

In an optical model, Q will typically be a three-dimensional Euclidean *luminiferous medium* \mathbb{E} , whose properties are characterized by the Hamiltonian $H = f_\Gamma$. If the

restriction $H_q = H|_{T_q^*\mathbb{E}} : T_q^*\mathbb{E} \rightarrow \mathbb{R}$ is the same (for instance, the same constant) for every point q of \mathbb{E} the medium is *homogeneous*, otherwise *inhomogeneous*; if H_q has spherical symmetry (with respect to the background Euclidean metric) on every cotangent space $T_q^*\mathbb{E}$ the medium is *isotropic*, otherwise *anisotropic*. Optical anisotropies tend to be of the ‘ellipsoidal’ kind represented by a tensor (field); the Hamiltonian will then be of the form $H = \langle p, g^\#(q)(p) \rangle$, where the tensor $g^\#(q)$ varies over the medium. But inhomogeneity and anisotropy are in principle independent, and can in some cases be neatly separated; when the same ellipsoidal anisotropy is common to the whole medium one can write $H = \langle p, \bar{g}^\#(p) \rangle + V(q)$, where the inhomogeneity is expressed by $V(q)$ and the common anisotropy by the constant tensor $\bar{g}^\#$. ‘Isotropies’ are represented by degeneracies (equal eigenvalues) of the tensor $g^\#$ (or $\bar{g}^\#$); if $g^\#$ is a multiple $g^\# = ge^\#$ of the background Euclidean metric $e^\#$ there will be complete isotropy and one can write $H = \bar{g}\langle p, e^\#(p) \rangle + V(q)$, where \bar{g} is a constant scalar.

The characteristic function $W = f_Q$ represents *time elapsed*, from when the luminous disturbance leaves the fundamental level surface Q_0 , here a *light source*. The level surfaces of W are *surfaces of equal time*. A point (q, p) on the cotangent bundle Γ indicates, in addition to a *position* $q \in Q$, a *normal slowness* (to use Hamilton’s term) $p = dW$, whose modulus is large where the surfaces of equal time are close together, small where they are far apart. So Γ is the set of all positions together with all normal slownesses. The modulus of the ‘contravariant slowness’ or ‘inverse velocity’ (neither term being Hamilton’s) $\dot{q} = p^\# = dH_q(p)$ is also large where the surfaces are close together. The congruence Φ_Γ represents the virtual set of all rays compatible with the properties of the medium, Φ_Q the more real set of all rays emanating orthogonally from the source Q_0 .

5 Mechanics

In mechanical models Q is a *configuration space* (determined by ideal holonomic scleronomic constraints), the cotangent bundle Γ *phase space*. A point $(q, p) \in \Gamma$ gives a *position* q together with a *momentum* p . The characteristic function W represents *action*, its level surfaces (such as Q_0) *surfaces of equal action*. The congruence Φ_Γ is the virtual set of all dynamical trajectories, for all possible initial conditions, generated by the Hamiltonian H , which represents *energy*. The smaller congruence Φ_Q is the set of trajectories crossing the level surface Q_0 orthogonally.

This remains rather abstract and general. We can restrict the models to a typical class involving point masses (‘PM’) by adding the basic objects P, m, V and axioms:

$$\text{Axiom 5: } P = \{\mathbf{p}_1, \dots, \mathbf{p}_N\}$$

$$\text{Axiom 6: } m : P \rightarrow \mathbb{R}^+ : \mathbf{p}_r \mapsto m_r = m(\mathbf{p}_r)$$

Axiom 7: $V : Q \rightarrow \mathbb{R}$

Axiom 8: $H(q, p) = \frac{1}{2} \langle p, g^\#(m_1, \dots, m_N)(p) \rangle + V.$

A system $\mathcal{H}' = (Q, \Gamma, f_M, \omega_M^\#, \Phi_M, Q_0, P, m, V)$ satisfying Axioms 1-8 will be a ‘PM model’ of the two-level Hamiltonian structure.

The mapping m assigns a positive and finite real number $m_r = m(\mathbf{p}_r)$ to each element \mathbf{p}_r of the finite nonempty set P . How the values m_1, \dots, m_N figure in the Hamiltonian requires comment. We begin with the empty three-dimensional Euclidean space \mathbb{E} , whose geometry is characterized by the Euclidean metric $e^\# : T_q^* \mathbb{E} \rightarrow T_q \mathbb{E}$ (all $q \in \mathbb{E}$). The metric $e_N^\# : T_q^* \mathbb{E}^N \rightarrow T_q \mathbb{E}^N$ of the N -th Cartesian power $\mathbb{E}^N = \mathbb{E} \times \dots \times \mathbb{E}$ remains Euclidean. The non-Euclidean manifold $Q \subset \mathbb{E}^N$ is an n -dimensional ($n \leq 3N$) subset of \mathbb{E}^N , and $T_q Q \subset T_q \mathbb{E}$ for all $q \in Q$. The non-Euclidean metric $e_Q^\# : T_q^* Q \rightarrow T_q Q$ agrees with $e_N^\#$ on all of the smaller domain $T_q^* Q$; it is the restriction $e_Q^\# = e_N^\#|_{T_q^* Q}$ of $e_N^\#$ to $T_q^* Q$.

To introduce m_1, \dots, m_N we first break up each cotangent space $T_q^* Q$ into N mutually orthogonal subspaces $T_q^* Q_1, \dots, T_q^* Q_N$, where $1 \leq \dim T_q^* Q_r \leq 3$, and

$$T_q^* Q = \bigoplus_{r=1}^N T_q^* Q_r.$$

The projection with range $T_q^* Q_r$ is denoted $\Pi_r : T_q^* Q \rightarrow T_q^* Q_r$. We will need the composition $\Pi_r^\# = e_Q^\# \circ \Pi_r : T_q^* Q \rightarrow T_q^* Q_r$, which ‘projects’ onto the subspace $T_q^* Q_r$. So each $\mathbf{p}_r \in P$ now has its own m_r , as well as a one-, two- or three-dimensional subspace $T_q^* Q_r \subset T_q^* Q$, and a $\Pi_r^\#$. Writing

$$g^\# = \sum_{r=1}^N m_r^{-1} \Pi_r^\#,$$

we have a metric which associates an m_r^{-1} with every subspace $T_q^* Q_r$; and the scalar product $\langle p, g^\#(p) \rangle$, central to the Hamiltonian, is based on $g^\#$. This explains how the values $m_r = m(\mathbf{p}_r)$ assigned to the elements $\mathbf{p}_r \in P$ enter into the Hamiltonian.

The elements of P are of course *particles*. The function m represents *mass*, and V *potential energy*—which, in the spirit of analytical mechanics, is included instead of *force*, dV .

6 Final Remarks

I have associated terms like ‘energy’ or ‘light source’ with the basic objects and left it at that, deliberately avoiding the separate and delicate issues of what such things really are and how they ought to be measured. McKinsey *et al.* (1953) expressed a similar attitude:

Thus we leave aside all problems connected with the measurement of time (as well as of mass, distance, and force); this is not because we believe all such problems have been solved—or that they are unimportant—but merely because we consider it is

possible to separate mechanics proper from such epistemological and experimental questions.

They continued in a footnote:

We may add that the attempt to deal with both sorts of problems simultaneously is in our opinion responsible for much of the confusion and murkiness characteristic of the usual discussions of the foundations of physics.

Truesdell replied with the following footnote in the same article:

The communicator is in complete disagreement with the view of classical mechanics expressed in this article. ... While he does not believe the present work achieves any progress whatever toward the precision of the concept of force, which always has been and remains still the central conceptual problem, and indeed the only one not essentially trivial, in the foundations of classical mechanics ...

The focus here has been, rather, on structural analogy and models.

Acknowledgements

I wish to thank Jeremy Butterfield, Sebastiano Carpi, Dennis Dieks, Hans Duistermaat, Remko Muis, Fred Muller, Michael Seevinck and Jos Uffink for many fruitful discussions.

References

- Arnol'd, V. I.: 1997, *Mathematical methods of classical mechanics*, Berlin, Springer
- Balzer, W., C. U. Moulines and J. D. Sneed: 1987, *An architectonic for science*, Dordrecht, Kluwer
- Broglie, L. de: 1925, Recherches sur la théorie des quanta, *Annales de physique* **3**, 22-128
- McKinsey, J. C. C., A. C. Sugar, and P. Suppes: 1953, Axiomatic foundations of classical particle mechanics, *Journal of rational mechanics and analysis* **2**, 253-72
- McKinsey, J. C. C. and P. Suppes: 1953, Transformations of Systems of Classical Particle Mechanics, *Journal of rational mechanics and analysis* **2**, 273-89
- Moulines, U. M. and J. D. Sneed: 1979, Suppes' philosophy of physics, pp.59-91 in R. Bogdan (editor) *Patrick Suppes*, Dordrecht, Reidel, 1979
- Schrödinger, E.: 1926, Quantisierung als Eigenwertproblem (zweite Mitteilung), *Annalen der Physik* **79**, 489-527
- Schrödinger, E.: 1934, Der Grundgedanke der Wellenmechanik, in Schrödinger, *Was ist ein Naturgesetz?*, Munich, Oldenbourg, 1987
- Sneed, J. D.: 1979, *The logical structure of mathematical physics*, Dordrecht, Kluwer
- Stegmüller, W.: 1973, *Theorie und Erfahrung*, Berlin, Springer
- Stegmüller, W.: 1980, *Neue Wege der Wissenschaftsphilosophie*, Berlin, Springer

Suppes, P.: 1999, *Introduction to logic*, New York, Dover

Suppes, P.: 2002, Representation and invariance of scientific structures, Stanford, CSLI

Wessel, L.: 1979, Schrödinger's route to wave mechanics, *Studies in history and philosophy of science*
10, 311-40